

FirstDIG

First Data Investigation on the Grid

<http://www.epcc.ed.ac.uk/firstdig/>

FirstDIG: Introduction

The FirstDIG project is a collaboration between First plc and the National e-Science Centre (represented by EPCC). The project aims to deploy an early implementation of the OGSA Data Access and Integration Services (OGSA-DAI) within the First South Yorkshire bus operational environment. The project has two central goals:

- 1) To demonstrate the deployment of OGSA-DAI services in a commercial environment, and learn from this process.
- 2) To answer specific business questions posed by First through a short data mining analysis using the OGSA-DAI service enabled data sources.

The project started in May 2003 and will run through to January 2004.

Background

The transformation of the Grid from an enabling technology in the scientific domain to a widely used business tool is a key requirement if the success expected by the UK Government is to be realised. To date the UK e-Science Programme has involved a large number of companies in collaborative projects.

Many of these companies are either IT related or classic early adopters. Few projects involve service companies such as First plc. However, the company has a clear business problem the Grid can help them solve and for this reason they are collaborating in this project.

OGSA, an extension of XML-based Web services, provides a clear entry point for many companies into the Grid domain. Data Access and Integration services are a key component of this.

Since its formation in 1995, First plc has striven to improve the opportunities for bus and rail travel in cities and towns across the UK.



Operating world-wide in many different transport sectors, the company runs over 10,000 vehicles in the UK and has 23% of the market, making it the UK's largest operator. First is represented in this project by First South Yorkshire buses.

Many businesses turn to data mining to better inform their decision-making. In the transport industry, the huge range of fragmented data sources has hindered its adoption.

In the FirstDIG project we intend to demonstrate how OGSA-DAI services can provide cost-effective access to disparate data sources and hence enable data mining of these sources to answer specific business questions.

•Data Sources

The data sources to be used in the project are from the following systems:

- Customer Contact – this records correspondence with customers including commendations and complaints.
- Vehicle Mileage – this records the daily vehicle mileage for bus services.
- Ticket Revenue – this contains the daily tickets sold and the money taken for the bus services.
- Schedule Adherence – a satellite tracking system that records whether a bus is arriving and departing on time from a bus stop.

These data sources range from SQL sources to ODBC sources to COBOL files and are located at various company sites.

It is precisely these issues that any technology must address in order to be applicable and useful to business.

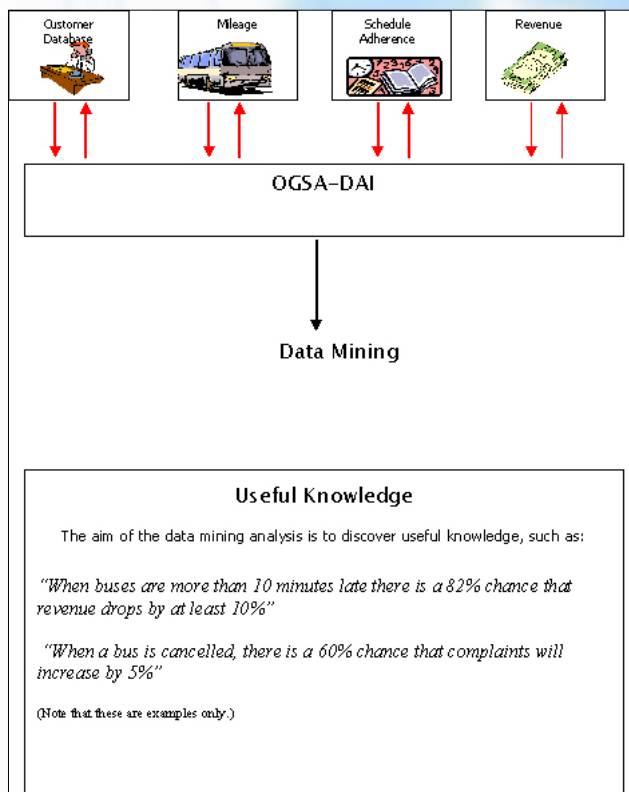
Data Mining and OGSA-DAI

Through data mining and statistical analyses the FirstDIG project aims to answer specific business questions posed by First.

This will require the consolidation of data from the customer contact system, the satellite tracking schedule adherence system, the mileage records system and the revenue system.

The project will determine if OGSA-DAI can be satisfactorily used to extract and consolidate the necessary data from these systems to answer these business questions.

(see figure, left)



Further information

For more information on the project and its deliverables please access the project web site at :

<http://www.epcc.ed.ac.uk/firstdig/>